

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開2000-47952

(P2000-47952A)

(43)公開日 平成12年2月18日(2000.2.18)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テ-マコード(参考)
G 0 6 F 13/00	3 5 1	G 0 6 F 13/00	3 5 1 E 5 B 0 8 2
12/00	5 1 4	12/00	5 1 4 A 5 B 0 8 9

審査請求 未請求 請求項の数8 O L (全 18 頁)

(21)出願番号 特願平10-211357

(22)出願日 平成10年7月27日(1998.7.27)

(71)出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(72)発明者 友田 正憲

東京都青梅市末広町2丁目9番地 株式会  
社東芝青梅工場内

(74)代理人 100081732

弁理士 大胡 典夫 (外1名)

Fターム(参考) 5B082 FA02 FA12

5B089 AA21 AA22 AC05 AD02 AD06

AE09 AF01 AF02 CA11 CB02

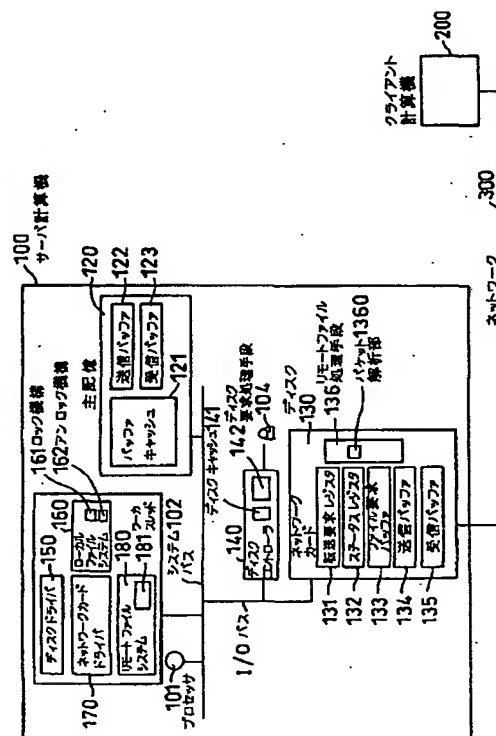
CB03 CB11

(54)【発明の名称】 ネットワークファイルサーバシステム、及び同システムに於けるファイル管理方法

(57)【要約】

【課題】 ネットワークアダプタやディスクコントローラ等のI/Oデバイス間でデータを直接やりとりすることで、サーバ計算機のプロセッサ、主記憶、システムバス資源の使用を最小限に抑え、効率良くI/O処理を行う手段を提供すること。

【解決手段】 サーバ計算機100上で動作するアプリケーション、OSがディスク上のファイルを、ローカルファイルシステム160を通して使用しつつ、ネットワークカード130がリモートファイル処理手段136により、リモートファイルシステム要求を処理する際、ディスクコントローラ140とネットワークカード130の間で直接データを転送することで、ディスクコントローラ140とネットワークカード130間の主記憶120を用いたデータ転送の回数を減らすことで、高速な処理を行える。



## 【特許請求の範囲】

【請求項1】 各種アプリケーション等のファイルを提供するサーバ計算機と、このサーバ計算機の提供するファイルにアクセスする複数のクライアント計算機と、上記サーバ計算機と上記クライアント計算機とを接続するネットワークとから成るネットワークファイルサーバシステムであって、

上記サーバ計算機は、

ファイル等の情報を格納する記憶装置と、

この記憶装置を制御する記憶装置制御手段と、

上記クライアント計算機に送受信するバケットのプロトコル処理をし、上記クライアント計算機へ送信するバケットを格納する送信バッファ部と上記クライアント計算機から受信したバケットを格納する受信バッファ部と上記クライアント計算機からのリモートファイルシステム要求に基づいた処理を行うリモートファイル処理部とを設けた通信手段と、

上記記憶装置を管理しこの記憶装置上のファイルを管理して、上記ファイルの上記記憶装置上の位置を問い合わせる要求に対して上記ファイルを格納しているディスクの位置を特定する情報を返す機能を持つローカルファイル制御手段と、

上記クライアント計算機からのリモートファイルシステム要求の内容に基づき上記記憶装置上のファイルを読み書きするリモートファイル制御手段とを備え、

上記リモートファイル処理部は、

上記受信バッファが受信したバケットを解析し、このバケットに含まれている上記クライアント計算機が送信した上記ローカルファイル制御手段の管理するファイルに対する操作要求であるリモートファイルシステム要求を取り出し、

このリモートファイルシステム要求が上記ローカルファイル制御手段の管理する上記記憶装置からのデータ転送を必要とする場合、上記ローカルファイル制御手段にそのファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記記憶装置制御手段に当該ファイルのデータを得られた位置情報に基づいて上記送信バッファに転送することを要求し、このデータが上記送信バッファに格納されたら送信バケットを作成して、上記クライアント計算機へリモートファイルシステム要求の返信を行い、

上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記受信バッファに格納しているデータを上記記憶装置制御手段に直接転送し、上記記憶装置制御手段が先に得た位置に基づき上記受信バッファに格納しているデータを上記記憶装置に書き込み、上記クライアント計算機へリモートファイルシステム要求に対する返信を行い、

上記リモートファイルシステム要求が上記記憶装置からのデータ転送を必要としない場合は、上記リモートファイル制御手段にその要求を渡すことを特徴とするネットワークファイルサーバシステム。

【請求項2】 各種アプリケーション等のファイルを提供するサーバ計算機と、このサーバ計算機の提供するファイルにアクセスする複数のクライアント計算機と、上記サーバ計算機と上記クライアント計算機とを接続するネットワークとから成るネットワークファイルサーバシステムであって、

上記サーバ計算機は、

ファイル等の情報を格納する記憶装置と、

この記憶装置を制御する記憶装置制御手段と、

上記クライアント計算機に送受信するバケットのプロトコル処理をし、上記クライアント計算機へ送信するバケットを格納する送信バッファ部と上記クライアント計算機から受信したバケットを格納する受信バッファ部と上記クライアント計算機からのリモートファイルシステム要求に基づいた処理を行うリモートファイル処理部とを設けた通信手段と、

上記ファイルのデータを上記サーバ計算機の主記憶上のバッファキャッシュ部に格納しておき、上記ファイルに対する操作は、上記バッファキャッシュ部上で行い、上記ファイルのデータをバッファキャッシュ部上に格納している場合には、このファイル要求に対して、上記バッファキャッシュ上の当該ファイルのデータのアドレスを返す機能と、上記記憶装置を管理しこの記憶装置上のファイルを管理して、上記ファイルの上記記憶装置上の位置を問い合わせる要求に対して上記ファイルを格納しているディスクの位置を特定する情報を返す機能とを設けたローカルファイル制御手段と、

上記クライアント計算機からのリモートファイルシステム要求の内容に基づき上記記憶装置上のファイルを読み書きするリモートファイル制御手段とを備え、

上記リモートファイル処理部は、

上記受信バッファが受信したバケットを解析し、このバケットに含まれている上記クライアント計算機が送信した上記ローカルファイル制御手段の管理するファイルに対する操作要求であるリモートファイルシステム要求を取り出し、

上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記バッファキャッシュ部のどの位置に存在するかを問い合わせる要求をし、

上記バッファキャッシュ部上に存在する場合は、当該ファイルのデータのアドレスに基づいてデータをバッファキャッシュ部から上記送信バッファ部にコピーし、当該データを含む送信バケットを作成して、上記クライアント計算機へリモートファイルシステム要求の返信を行い、

上記バッファキャッシュ部に存在しない場合は、上記ローカルファイル制御手段にそのファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記記憶装置制御手段に当該ファイルのデータを得られた位置情報に基づいて上記送信バッファに転送することを要求し、このデータが上記送信バッファに格納されたら送信バケットを作成して、上記クライアント計算機ヘリモートファイルシステム要求の返信を行い、

上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記バッファキャッシュ部のどの位置に存在するかを問い合わせる要求をし、

上記バッファキャッシュ部に存在する場合は、当該ファイルのデータのアドレスに基づいて上記受信バッファのデータを書き込み、上記クライアント計算機ヘリモートファイルシステム要求に対する返信を行い、

上記バッファキャッシュ部に存在しない場合は、上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記受信バッファに格納しているデータを上記記憶装置制御手段に直接転送し、上記記憶装置制御手段が先に得た位置に基づき上記受信バッファに格納しているデータを上記記憶装置に書き込み、上記クライアント計算機ヘリモートファイルシステム要求に対する返信を行い、

リモートファイルシステム要求が上記記憶装置からのデータ転送を必要としない場合は、上記リモートファイル制御手段にその要求を渡すことを特徴とするネットワークファイルサーバシステム。

【請求項3】 各種アプリケーション等のファイルを提供するサーバ計算機と、このサーバ計算機の提供するファイルにアクセスする複数のクライアント計算機と、上記サーバ計算機と上記クライアント計算機とを接続するネットワークとから成るネットワークファイルサーバシステムであって、

上記サーバ計算機は、

ファイル等の情報を格納する記憶装置と、

この記憶装置を制御する記憶装置制御手段と、

上記クライアント計算機に送受信するバケットのプロトコル処理をし、上記クライアント計算機へ送信するバケットを格納する送信バッファ部と上記クライアント計算機から受信したバケットを格納する受信バッファ部と上記クライアント計算機からのリモートファイルシステム要求に基づいた処理を行うリモートファイル処理部とを設けた通信手段と、

上記ファイルのデータを上記サーバ計算機の主記憶上のバッファキャッシュ部に格納しておき、上記ファイルに対する操作は、上記バッファキャッシュ部上で行い、上記ファイルのデータをバッファキャッシュ部上に格納し

ている場合には、このファイル要求に対して、上記バッファキャッシュ上の当該ファイルのデータのアドレスを返す機能と、上記ファイルのデータをバッファキャッシュ部に格納していない場合、上記ファイル要求に基づいて、当該ファイルを格納する上記バッファキャッシュ部を確保し、このアドレスを返す機能と、上記記憶装置を管理しこの記憶装置上のファイルを管理して、上記ファイルの上記記憶装置上の位置を問い合わせる要求に対して上記ファイルを格納しているディスクの位置を特定する情報を返す機能とを設けたローカルファイル制御手段と、

上記クライアント計算機からのリモートファイルシステム要求の内容に基づき上記記憶装置上のファイルを読み書きするリモートファイル制御手段とを備え、上記リモートファイル処理部は、

上記受信バッファが受信したバケットを解析し、このバケットに含まれている上記クライアント計算機が送信した上記ローカルファイル制御手段の管理するファイルに対する操作要求であるリモートファイルシステム要求を取り出し、

上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記バッファキャッシュ部のどの位置に存在するかを問い合わせる要求をし、

上記バッファキャッシュ部に存在する場合は、当該ファイルのデータのアドレスに基づいてデータをバッファキャッシュ部から上記送信バッファ部にコピーし、当該データを含む送信バケットを作成して、上記クライアント計算機ヘリモートファイルシステム要求の返信を行い、

上記バッファキャッシュ部に存在しない場合は、上記ローカルファイル制御手段にそのファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記記憶装置制御手段に当該ファイルのデータを得られた位置情報に基づいて上記送信バッファに転送することを要求し、このデータが上記送信バッファに格納されたら送信バケットを作成して、上記クライアント計算機ヘリモートファイルシステム要求の返信を行い、上記ローカルファイル制御手段が確保したバッファキャッシュ部のアドレスに上記データを転送し、

上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記バッファキャッシュ部のどの位置に存在するかを問い合わせる要求をし、

上記バッファキャッシュ部に存在する場合は、当該ファイルのデータのアドレスに基づいて上記受信バッファのデータを書き込み、上記クライアント計算機ヘリモートファイルシステム要求に対する返信を行い、

上記バッファキャッシュ部に存在しない場合は、上記リモートファイルシステム要求が上記記憶装置へのデー

タ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記受信バッファに格納しているデータを上記記憶装置制御手段に直接転送し、上記記憶装置制御手段が先に得た位置に基づき上記受信バッファに格納しているデータを上記記憶装置に書き込み、上記クライアント計算機ヘリモートファイルシステム要求に対する返信を行い、上記ローカルファイル制御手段が確保したバッファキャッシュ部のアドレスに上記受信バッファに格納しているデータを転送し、リモートファイルシステム要求が上記記憶装置からのデータ転送を必要としない場合は、上記リモートファイル制御手段にその要求を渡すことを特徴とするネットワークファイルサーバシステム。

【請求項4】 各種アプリケーション等のファイルを提供するサーバ計算機と、このサーバ計算機の提供するファイルにアクセスする複数のクライアント計算機と、上記サーバ計算機と上記クライアント計算機とを接続するネットワークとから成るネットワークファイルサーバシステムであって、

上記サーバ計算機は、

ファイル等の情報を格納する記憶装置と、

この記憶装置を制御する記憶装置制御手段と、

上記クライアント計算機に送受信するパケットのプロトコル処理をし、上記クライアント計算機へ送信するパケットを格納する送信バッファ部と上記クライアント計算機から受信したパケットを格納する受信バッファ部とを設け、上記クライアント計算機からのリモートファイルシステム要求に基づき処理を行うリモートファイル処理部とを設けた通信手段と、

上記記憶装置を管理しこの記憶装置上のファイルを管理して、上記ファイルの上記記憶装置上の位置を問い合わせる要求に対して上記ファイルを格納しているディスクの位置を特定する情報を返す機能と、上記ファイルへの書き込み、削除、作成等の操作を行う際に、このファイルを格納する記憶装置上の位置情報とこのファイルを特定する情報であるファイル位置情報を上記通信手段に通知する機能とを備えたローカルファイル制御手段と、

上記クライアント計算機からのリモートファイルシステム要求の内容に基づき上記記憶装置上のファイルを読み書きするリモートファイル制御手段とを備え、上記リモートファイル処理部は、

上記受信バッファが受信したパケットを解析し、このパケットに含まれている上記クライアント計算機が送信した上記ローカルファイル制御手段の管理するファイルに対する操作要求であるリモートファイルシステム要求を取り出し、

このリモートファイルシステム要求が上記ローカルファイル制御手段の管理する上記記憶装置からのデータ転送を必要とする場合、上記ファイル位置情報に基づき一致

するファイルがあるかを検索し、一致した場合、上記記憶装置制御手段に上記ファイル位置情報に基づき上記記憶装置上のデータを上記送信バッファに転送することを要求し、このデータが上記送信バッファに格納されたら送信パケットを作成して、上記クライアント計算機ヘリモートファイルシステム要求の返信を行い、

一致しない場合、上記ローカルファイル制御手段にそのファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記記憶装置制御手段に当該ファイルのデータを得られた位置情報に基づいて上記送信バッファに転送することを要求し、このデータが上記送信バッファに格納されたら送信パケットを作成して、上記クライアント計算機ヘリモートファイルシステム要求の返信を行い、

上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ファイル位置情報に基づき一致するファイルがあるかを検索し、一致した場合、上記ファイル位置情報に基づき上記受信バッファに格納しているデータを上記記憶装置制御手段に転送し、上記記憶装置制御手段は上記ファイル位置情報に基づき上記受信バッファに格納しているデータを上記記憶装置上に書き込み、上記クライアント計算機ヘリモートファイルシステム要求に対する返信を行い、

一致しない場合、上記ローカルファイル制御手段に当該ファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記受信バッファに格納しているデータを上記記憶装置制御手段に直接転送し、上記記憶装置制御手段が先に得た位置に基づき上記受信バッファに格納しているデータを書き込み、上記クライアント計算機ヘリモートファイルシステム要求の返信を行い、リモートファイルシステム要求が上記記憶装置からのデータ転送を必要としない場合は、上記リモートファイル制御手段にその要求を渡すことを特徴とするネットワークファイルサーバシステム。

【請求項5】 各種アプリケーション等のファイルを提供するサーバ計算機と、このサーバ計算機の提供するファイルにアクセスする複数のクライアント計算機と、上記サーバ計算機と上記クライアント計算機とを接続するネットワークとから成るネットワークファイルサーバシステムに於けるファイル管理方法であって、

上記サーバ計算機は、

ファイル等の情報を格納する記憶装置と、

この記憶装置を制御する記憶装置制御手段と、

上記クライアント計算機に送受信するパケットのプロトコル処理をし、上記クライアント計算機へ送信するパケットを格納する送信バッファ部と上記クライアント計算機から受信したパケットを格納する受信バッファ部と上記クライアント計算機からのリモートファイルシステム要求に基づいた処理を行うリモートファイル処理部とを設けた通信手段と、

上記記憶装置を管理しこの記憶装置上のファイルを管理して、上記ファイルの上記記憶装置上の位置を問い合わせる要求に対して上記ファイルを格納しているディスクの位置を特定する情報を返す機能を持つローカルファイル制御手段と、

上記クライアント計算機からのリモートファイルシステム要求の内容に基づき上記記憶装置上のファイルを読み書きするリモートファイル制御手段とを備え、

上記リモートファイル処理部は、

上記受信バッファが受信したパケットを解析し、このパケットに含まれている上記クライアント計算機が送信した上記ローカルファイル制御手段の管理するファイルに対する操作要求であるリモートファイルシステム要求を取り出し、

このリモートファイルシステム要求が上記ローカルファイル制御手段の管理する上記記憶装置からのデータ転送を必要とする場合、上記ローカルファイル制御手段にそのファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記記憶装置制御手段に当該ファイルのデータを得られた位置情報に基づいて上記送信バッファに転送することを要求し、このデータが上記送信バッファに格納されたら送信パケットを作成して、上記クライアント計算機へリモートファイルシステム要求の返信を行い、

上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記受信バッファに格納しているデータを上記記憶装置制御手段に直接転送し、上記記憶装置制御手段が先に得た位置に基づき上記受信バッファに格納しているデータを上記記憶装置に書き込み、上記クライアント計算機へリモートファイルシステム要求に対する返信を行い、

上記リモートファイルシステム要求が上記記憶装置からのデータ転送を必要としない場合は、上記リモートファイル制御手段にその要求を渡すことを特徴とするネットワークファイルサーバシステムに於けるファイル管理方法。

【請求項6】 各種アプリケーション等のファイルを提供するサーバ計算機と、このサーバ計算機の提供するファイルにアクセスする複数のクライアント計算機と、上記サーバ計算機と上記クライアント計算機とを接続するネットワークとからなるネットワークファイルサーバシステムのファイル管理方法であって、

上記サーバ計算機は、

ファイル等の情報を格納する記憶装置と、

この記憶装置を制御する記憶装置制御手段と、

上記クライアント計算機に送受信するパケットのプロトコル処理をし、上記クライアント計算機へ送信するパケットを格納する送信バッファ部と上記クライアント計算

機から受信したパケットを格納する受信バッファ部と上記クライアント計算機からのリモートファイルシステム要求に基づいた処理を行うリモートファイル処理部とを設けた通信手段と、

上記ファイルのデータを上記サーバ計算機の主記憶上のバッファキャッシュ部に格納しておき、上記ファイルに対する操作は、上記バッファキャッシュ部上で行い、上記ファイルのデータをバッファキャッシュ部上に格納している場合には、このファイル要求に対して、上記バッファキャッシュ上の当該ファイルのデータのアドレスを返す機能と、上記記憶装置を管理しこの記憶装置上のファイルを管理して、上記ファイルの上記記憶装置上の位置を問い合わせる要求に対して上記ファイルを格納しているディスクの位置を特定する情報を返す機能とを設けたローカルファイル制御手段と、

上記クライアント計算機からのリモートファイルシステム要求の内容に基づき上記記憶装置上のファイルを読み書きするリモートファイル制御手段とを備え、上記リモートファイル処理部は、

上記受信バッファが受信したパケットを解析し、このパケットに含まれている上記クライアント計算機が送信した上記ローカルファイル制御手段の管理するファイルに対する操作要求であるリモートファイルシステム要求を取り出し、

上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記バッファキャッシュ部のどの位置に存在するかを問い合わせる要求をし、

上記バッファキャッシュ部上に存在する場合は、当該ファイルのデータのアドレスに基づいてデータをバッファキャッシュ部から上記送信バッファ部にコピーし、当該データを含む送信パケットを作成して、上記クライアント計算機へリモートファイルシステム要求の返信を行い、

上記バッファキャッシュ部に存在しない場合は、上記ローカルファイル制御手段にそのファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記記憶装置制御手段に当該ファイルのデータを得られた位置情報に基づいて上記送信バッファに転送することを要求し、このデータが上記送信バッファに格納されたら送信パケットを作成して、上記クライアント計算機へリモートファイルシステム要求の返信を行い、

上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記バッファキャッシュ部のどの位置に存在するかを問い合わせる要求をし、

上記バッファキャッシュ部上に存在する場合は、当該ファイルのデータのアドレスに基づいて上記受信バッファのデータを書き込み、上記クライアント計算機へリモートファイルシステム要求に対する返信を行い、

上記バッファキャッシュ部に存在しない場合は、上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記受信バッファに格納しているデータを上記記憶装置制御手段に直接転送し、上記記憶装置制御手段が先に得た位置に基づき上記受信バッファに格納しているデータを上記記憶装置に書き込み、上記クライアント計算機へリモートファイルシステム要求に対する返信を行い、

リモートファイルシステム要求が上記記憶装置からのデータ転送を必要としない場合は、上記リモートファイル制御手段にその要求を渡すことを特徴とするネットワークファイルサーバシステムに於けるファイル管理方法。

【請求項7】 各種アプリケーション等のファイルを提供するサーバ計算機と、このサーバ計算機の提供するファイルにアクセスする複数のクライアント計算機と、上記サーバ計算機と上記クライアント計算機とを接続するネットワークとからなるネットワークファイルサーバシステムのファイル管理方法であって、

上記サーバ計算機は、

ファイル等の情報を格納する記憶装置と、

この記憶装置を制御する記憶装置制御手段と、

上記クライアント計算機に送受信するパケットのプロトコル処理をし、上記クライアント計算機へ送信するパケットを格納する送信バッファ部と上記クライアント計算機から受信したパケットを格納する受信バッファ部と上記クライアント計算機からのリモートファイルシステム要求に基づいた処理を行うリモートファイル処理部とを設けた通信手段と、

上記ファイルのデータを上記サーバ計算機の主記憶上のバッファキャッシュ部に格納しておき、上記ファイルに対する操作は、上記バッファキャッシュ部上で行い、上記ファイルのデータをバッファキャッシュ部に格納している場合には、このファイル要求に対して、上記バッファキャッシュ上の当該ファイルのデータのアドレスを返す機能と、上記ファイルのデータをバッファキャッシュ部に格納していない場合、上記ファイル要求に基づいて、当該ファイルを格納する上記バッファキャッシュ部を確保し、このアドレスを返す機能と、上記記憶装置を管理しこの記憶装置上のファイルを管理して、上記ファイルの上記記憶装置上の位置を問い合わせる要求に対して上記ファイルを格納しているディスクの位置を特定する情報を返す機能とを設けたローカルファイル制御手段と、

上記クライアント計算機からのリモートファイルシステム要求の内容に基づき上記記憶装置上のファイルを読み書きするリモートファイル制御手段とを備え、

上記リモートファイル処理部は、

上記受信バッファが受信したパケットを解析し、このパ

ケットに含まれている上記クライアント計算機が送信した上記ローカルファイル制御手段の管理するファイルに対する操作要求であるリモートファイルシステム要求を取り出し、

上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記バッファキャッシュ部のどの位置に存在するかを問い合わせる要求をし、

上記バッファキャッシュ部に存在する場合は、当該ファイルのデータのアドレスに基づいてデータをバッファキャッシュ部から上記送信バッファ部にコピーし、当該データを含む送信パケットを作成して、上記クライアント計算機へリモートファイルシステム要求の返信を行い、

上記バッファキャッシュ部に存在しない場合は、上記ローカルファイル制御手段にそのファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記記憶装置制御手段に当該ファイルのデータを得られた位置情報に基づいて上記送信バッファに転送することを要求し、このデータが上記送信バッファに格納されたら送信パケットを作成して、上記クライアント計算機へリモートファイルシステム要求の返信を行い、上記ローカルファイル制御手段が確保したバッファキャッシュ部のアドレスに上記データを転送し、

上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記バッファキャッシュ部のどの位置に存在するかを問い合わせる要求をし、

上記バッファキャッシュ部に存在する場合は、当該ファイルのデータのアドレスに基づいて上記受信バッファのデータを書き込み、上記クライアント計算機へリモートファイルシステム要求に対する返信を行い、

上記バッファキャッシュ部に存在しない場合は、上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記受信バッファに格納しているデータを上記記憶装置制御手段に直接転送し、上記記憶装置制御手段が先に得た位置に基づき上記受信バッファに格納しているデータを上記記憶装置に書き込み、上記クライアント計算機へリモートファイルシステム要求に対する返信を行い、上記ローカルファイル制御手段が確保したバッファキャッシュ部のアドレスに上記受信バッファに格納しているデータを転送し、

リモートファイルシステム要求が上記記憶装置からのデータ転送を必要としない場合は、上記リモートファイル制御手段にその要求を渡すことを特徴とするネットワークファイルサーバシステムに於けるファイル管理方法。

【請求項8】 各種アプリケーション等のファイルを提供するサーバ計算機と、このサーバ計算機の提供するフ



ファイルにアクセスする複数のクライアント計算機と、上記サーバ計算機と上記クライアント計算機とを接続するネットワークとからなるネットワークファイルサーバシステムのファイル管理方法であって、

上記サーバ計算機は、

ファイル等の情報を格納する記憶装置と、

この記憶装置を制御する記憶装置制御手段と、

上記クライアント計算機に送受信するバケットのプロトコル処理をし、上記クライアント計算機へ送信するバケットを格納する送信バッファ部と上記クライアント計算機から受信したバケットを格納する受信バッファ部とを設け、上記クライアント計算機からのリモートファイルシステム要求に基づき処理を行うリモートファイル処理部とを設けた通信手段と、

上記記憶装置を管理しこの記憶装置上のファイルを管理して、上記ファイルの上記記憶装置上の位置を問い合わせる要求に対して上記ファイルを格納しているディスクの位置を特定する情報を返す機能と、上記ファイルへの書き込み、削除、作成等の操作を行う際に、このファイルを格納する記憶装置上の位置情報とこのファイルを特定する情報であるファイル位置情報を上記通信手段に通知する機能とを備えたローカルファイル制御手段と、上記クライアント計算機からのリモートファイルシステム要求の内容に基づき上記記憶装置上のファイルを読み書きするリモートファイル制御手段とを備え、上記リモートファイル処理部は、

上記受信バッファが受信したバケットを解析し、このバケットに含まれている上記クライアント計算機が送信した上記ローカルファイル制御手段の管理するファイルに対する操作要求であるリモートファイルシステム要求を取り出し、

このリモートファイルシステム要求が上記ローカルファイル制御手段の管理する上記記憶装置からのデータ転送を必要とする場合、上記ファイル位置情報に基づき一致するファイルがあるかを検索し、一致した場合、上記記憶装置制御手段に上記ファイル位置情報に基づき上記記憶装置上のデータを上記送信バッファに転送することを要求し、このデータが上記送信バッファに格納されたら送信バケットを作成して、上記クライアント計算機へリモートファイルシステム要求の返信を行い、

一致しない場合、上記ローカルファイル制御手段にそのファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記記憶装置制御手段に当該ファイルのデータを得られた位置情報に基づいて上記送信バッファに転送することを要求し、このデータが上記送信バッファに格納されたら送信バケットを作成して、上記クライアント計算機へリモートファイルシステム要求の返信を行い、

上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ファイル位置情報に

基づき一致するファイルがあるかを検索し、一致した場合、上記ファイル位置情報に基づき上記受信バッファに格納しているデータを上記記憶装置制御手段に転送し、上記記憶装置制御手段は上記ファイル位置情報に基づき上記受信バッファに格納しているデータを上記記憶装置上に書き込み、上記クライアント計算機へリモートファイルシステム要求に対する返信を行い、一致しない場合、上記ローカルファイル制御手段に当該ファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記受信バッファに格納しているデータを上記記憶装置制御手段に直接転送し、上記記憶装置制御手段が先に得た位置に基づき上記受信バッファに格納しているデータを書き込み、上記クライアント計算機へリモートファイルシステム要求の返信を行い、リモートファイルシステム要求が上記記憶装置からのデータ転送を必要としない場合は、上記リモートファイル制御手段にその要求を渡すことを特徴とするネットワークファイルサーバシステムに於けるファイル管理方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、複数のクライアント計算機から公衆通信網や有線、無線のLAN (Local Area Network) 等のネットワークを介してサーバ計算機上のファイルにアクセスし、データの送受信を行うネットワークファイルサーバシステム、及びネットワークファイルサーバシステムに於けるファイル管理方法に関する。

【0002】

【従来の技術】従来、図8に示すように、ネットワークファイルサーバ10では、クライアント計算機25とネットワークファイルサーバ10の間で情報をやり取りするものであった。即ち、

(1) クライアント計算機25から、LAN等のネットワーク24を経由し、ファイル操作要求を格納したバケットがネットワークファイルサーバ10 (以下、サーバ計算機10と呼ぶ) のネットワークカード18に届く。

【0003】(2) ネットワークドライバ20が、サーバ計算機10のネットワークカード18に届いたバケットをネットワークカード18の受信バッファ19から、主記憶14に転送する。

【0004】(3) TCP/IP等の通信プロトコルを解析するプロトコルスタック21が、バケットの内容を解析し、バケットに格納されたファイル操作要求を取り出して、リモートファイルシステム22に渡す。

【0005】(4) リモートファイルシステム22は、ファイル操作要求をローカルファイルシステム23に渡す。ローカルファイルシステム23は、サーバ計算機10のディスク17を管理するファイルシステムである。

【0006】(5) ローカルファイルシステム23がファイル操作要求を処理する。その結果をリモートファイ

ルシステム22に返す。

【0007】(6)リモートファイルシステム22は、結果を格納したパケットを作成し、プロトコルスタック21、ネットワークカード18、ネットワーク24を経由して、結果をクライアント計算機25に送信する。

【0008】到着したパケットを処理するソフトウェアは、ネットワークドライバ20、プロトコルスタック21、リモートファイルシステム22、ローカルファイルシステム23である。これらは、ディスク等の記憶装置上にインストールされているものとする。

【0009】サーバ計算機10のプロセッサ11が、主記憶14上でこれらのソフトウェアを実行する。ファイル操作要求の処理では、主記憶14を経由して、ディスク17とネットワークカード18間のデータのやり取りを行う。

【0010】これまでのサーバ計算機10に於いて、信頼性の高いデータ送受信を実現するプロトコルスタック21の処理に時間がかかり、プロセッサ資源を多く消費していた。

【0011】また、ネットワークファイルサービス等に於いては、ディスクコントローラ16、ネットワークカード18を用いてデータのやり取りが発生する。どちらのデバイスも、サーバ計算機10のI/Oバス13（PCIバス等で実現する）に接続されている。

【0012】そのために、これまででは、ネットワークカード18、ディスクコントローラ16等のデバイスを制御するソフトウェア（サーバ計算機のプロセッサが実行する）が、データを主記憶にいったん転送し、処理を行った後、各デバイスへデータをさらに転送していたものであった。

【0013】

【発明が解決しようとする課題】上述したように、従来のサーバ計算機10にあっては、信頼性の高いデータ送受信を実現するプロトコルスタック21の処理に時間がかかり、プロセッサ資源を多く消費していた。また、ネットワークファイルサービス等に於いては、ディスクコントローラ16、ネットワークカード18を用いてデータのやり取りが発生するものであった。どちらのデバイスも、サーバ計算機10のI/Oバス13（PCIバス等で実現する）に接続されている。

【0014】そのために、従来は、ネットワークカード18、ディスクコントローラ16等のデバイスを制御するソフトウェア（サーバ計算機のプロセッサが実行する）が、データを主記憶にいったん転送し、処理を行った後、各デバイスへデータを更に転送していたものであった。

【0015】そこで、本発明は上記事情を考慮して成されたもので、上記不具合を解消し、複数のクライアント計算機から公衆通信網や有線、無線のLAN等のネットワークを介してサーバ計算機上のファイルにアクセス

し、データの送受信を行うネットワークファイルサーバシステムにあって、ネットワークアダプタやディスクコントローラ等のI/Oデバイス間でデータを直接やり取りすることで、サーバ計算機のプロセッサ、主記憶、システムバス資源の使用を最小限に抑え、効率良くI/O処理を行えるネットワークファイルサーバシステム、及びネットワークファイルサーバシステムに於けるファイル管理方法を提供することを目的とする。

【0016】

【課題を解決するための手段】本発明は上記目的を達成する為、本発明のネットワークファイルサーバシステムは、各種アプリケーション等のファイルを提供するサーバ計算機と、このサーバ計算機の提供するファイルにアクセスする複数のクライアント計算機と、上記サーバ計算機と上記クライアント計算機とを接続するネットワークとからなるネットワークファイルサーバシステムであって、上記サーバ計算機は、ファイル等の情報を格納する記憶装置と、この記憶装置を制御する記憶装置制御手段と、上記クライアント計算機に送受信するパケットのプロトコル処理をし、上記クライアント計算機へ送信するパケットを格納する送信バッファ部と上記クライアント計算機から受信したパケットを格納する受信バッファ部と上記クライアント計算機からのリモートファイルシステム要求に基づいた処理を行うリモートファイル処理部とを設けた通信手段と、上記記憶装置を管理しこの記憶装置上のファイルを管理して、上記ファイルの上記記憶装置上の位置を問い合わせる要求に対して上記ファイルを格納しているディスクの位置を特定する情報を返す機能を持つローカルファイル制御手段と、上記クライアント計算機からのリモートファイルシステム要求の内容に基づき上記記憶装置上のファイルを読み書きするリモートファイル制御手段とを備え、上記リモートファイル処理部は、上記受信バッファが受信したパケットを解析し、このパケットに含まれている上記クライアント計算機が送信した上記ローカルファイル制御手段の管理するファイルに対する操作要求であるリモートファイルシステム要求を取り出し、このリモートファイルシステム要求が上記ローカルファイル制御手段の管理する上記記憶装置からのデータ転送を必要とする場合、上記ローカルファイル制御手段にそのファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記記憶装置制御手段に当該ファイルのデータを得られた位置情報に基づいて上記送信バッファに転送することを要求し、このデータが上記送信バッファに格納されたら送信パケットを作成して、上記クライアント計算機へリモートファイルシステム要求の返信を行い、上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記記憶装置制御手段にファイルのデータ



を得られた位置情報に基づいて上記送信バッファに転送することを要求し、このデータが上記送信バッファに格納されたら送信パケットを作成して上記クライアント計算機へリモートファイルシステム要求の返信を行い、上記リモートファイルシステム要求が上記記憶装置へのデータ転送を必要とする場合、上記ローカルファイル制御手段に当該ファイルが上記記憶装置のどの位置に存在するかを問い合わせる要求をし、上記受信バッファに格納しているデータを上記記憶装置制御手段に直接転送し、上記記憶装置制御手段が先に得た位置に基づき上記受信バッファに格納しているデータを上記記憶装置に書き込み、上記クライアント計算機へリモートファイルシステム要求に対する返信を行い、上記リモートファイルシステム要求が上記記憶装置からのデータ転送を必要としない場合は、上記リモートファイル制御手段にその要求を渡すように構成したことを特徴とする。

【0017】このような構成によれば、サーバ計算機上で動作するアプリケーション、オペレーティングシステム（OS）がディスク上のファイルを、ローカルファイルシステムを通して使用しつつ、ネットワークカードが、リモートファイルシステム要求を処理する際に、ディスクコントローラとネットワークカードの間で直接データを転送することで、ディスクコントローラとネットワークカード間の主記憶を用いたデータ転送の回数を減らすことで、高速な処理を行うことができる。

【0018】さらに、ネットワークカードが解釈したパケットのうち、ネットワークファイルシステム要求のみを、本発明で提案するリモートファイルシステム（サーバ計算機のプロセッサが実行する）に送る。このとき、要求のみをサーバ計算機の主記憶に送り、ファイルのデータ等をネットワークカードとディスクコントローラの間で直接DMA転送することにより、これまでよりデータ転送の回数を減らすことができる。

【0019】この構成に加えて、上記ローカルファイル制御手段が、要求のあったファイルを格納する記憶装置上の位置情報とこのファイルを特定する情報であるファイル位置情報を上記通信手段に通知する機能を備えれば、更に、高速にファイルのディスク位置情報を得ることができる。

【0020】また、この構成に加えて、サーバ計算機上で動作するアプリケーション、OSがディスク上のファイルを、ローカルファイルシステムを通して使用しつつ、ネットワークカードが、リモートファイルシステム要求を処理する際に、ローカルファイルシステムが管理する主記憶上のバッファキャッシュのデータをネットワークカードに直接転送することで、高速な処理を行うことができ、更に、リモートファイルシステム要求で発生したディスクとのデータのやりとりをバッファキャッシュに格納することで、次に同じファイルへの要求（サーバ計算機上のアプリケーションがローカルファイルシス

テムを通して、あるいは、クライアント計算機がリモートファイルシステム要求として）に対し、バッファキャッシュ上のデータを使用することができる。

【0021】

【発明の実施の形態】以下、本発明の一実施の形態について、図面を参照して説明する。

【0022】図1は、本実施形態に係わるネットワークファイルサーバシステムの構成を示したブロック図である。本実施形態の説明に必要な構成を示し、他の構成は省略している。

【0023】クライアント計算機200は、公衆通信網や有線、無線のLAN等のネットワーク300によってサーバ計算機100と接続されており、サーバ計算機100上のファイルにアクセスする。このファイルは、サーバ計算機上にインストールされた各種アプリケーション、またはその他のアプリケーション等に対応したファイルである。

【0024】プロセッサ101は、サーバ計算機100の各部の制御を行うものである。

【0025】システムバス102は、主記憶120等とプロセッサ101を接続し各種データ転送するものである。

【0026】I/Oバス103は、ネットワークカード130、ディスクコントローラ140、図示しないキーボード、マウス等の入力手段、ディスプレイ等のI/O機器とシステムバス102を接続し各種データを転送する。

【0027】主記憶120は、サーバ計算機100の主記憶であり、バッファキャッシュ121、送信バッファ122、受信バッファ123を備える。

【0028】リモートファイルシステム180、ローカルファイルシステム160、ネットワークカードドライバ170、ディスクドライバ150は、ソフトウェアであり、ディスク104等の記憶装置上に保存され、プロセッサ101により実行される。

【0029】リモートファイルシステム180は、ワークスレッド181によりファイルシステム要求の内容に基づきサーバ計算機100上のファイルを読み書きする等の処理を行う。

【0030】ローカルファイルシステム160は、サーバ計算機100が備えるディスク104を管理し、ロック機能161とアンロック機能162を備える。ロック機能161とアンロック機能162の機能については後述する。

【0031】ネットワークカードドライバ170は、ネットワークカード130を制御するものである。

【0032】ディスクドライバ150は、ディスクコントローラ140によりディスク104を制御する。

【0033】サーバ計算機100は、クライアント計算機200とデータの送受信を行うためのプロトコル処理

を行うネットワークカード130を備える。このネットワークカード130は、サーバ計算機100とクライアント計算機200間で送受信するバケットのプロトコル処理を行った結果、そのバケットがクライアント計算機200からサーバ計算機100上のファイルを読み書き等するファイルシステム要求でない場合は、ネットワークカードドライバ170と連携して、通常のネットワークカードと同様の処理を行う。

【0034】また、ネットワークカード130は、上記送受信するバケットがファイルシステム要求の場合、サーバ計算機100上で動作するネットワークカードドライバ170と、リモートファイルシステム180上のワーカーズレッド181と協調して、ファイルシステム要求の内容に基づきサーバ計算機100上のファイルを読み書きする等の処理を行う。

【0035】上記ネットワークカード130は、以下のものから構成される。

【0036】受信バケットのプロトコル処理を実行し、その受信バケットがファイルシステム要求の場合には、その処理を行うリモートファイル処理手段136と、リモートファイル処理手段136が解析したバケットの内容を格納するファイル要求バッファ133、受信バケットが通常のバケットか、或いはファイルシステム要求か等を示す情報を格納するステータスレジスタ132、他の計算機から送信されたバケットをいったん格納する受信バッファ135と、送信すべきバケットを蓄積する送信バッファ134と転送要求レジスタ131とからなる。

【0037】転送要求レジスタ131は、データの格納元のアドレスとそこからデータ転送し、そのデータを格納する先のアドレスの情報が格納される。例えば、送信データがバッファキャッシュ121上にある場合は、そのデータのバッファキャッシュ121上のアドレスと、そのデータの転送先である送信バッファ134上のアドレスが格納される。受信データが受信バッファ135上にある場合、そのデータの受信バッファ135上のアドレスと、そのデータの転送先であるバッファキャッシュ121上のアドレスが格納される。

【0038】ステータスレジスタ132に格納するステータスを示す値は、ここでは次の通りとする。1はファイルシステムのリードライト要求受信を示し、2は通常バケット受信を示し、3はDMA転送完了を示し、そして、4は通常のバケット送信処理を示す。

【0039】ファイル要求バッファ133は、図2に示すようにリモートファイルシステム要求1330（リード、ライト、その他）、対象となるファイル名1331、そのオフセット1332、データ長1333、データアドレス1334から成るものである。

【0040】ローカルファイルシステム160は、サーバ計算機100が備えるディスクを管理する。即ち、デ

ィスク104上にファイルを格納し、ファイルの作成、リード、ライト、削除等の機能を提供する。主記憶120上にバッファキャッシュ121を作成し、ファイルのデータをキャッシュする。更に、以下の機能を備えるものである。

【0041】1. ファイル名、オフセット、データ長を入力として、そのファイルの該当する内容がバッファキャッシュ121にあるかどうかを判定する。バッファキャッシュ121にある場合には、そのバッファのアドレスを返す。

【0042】2. バッファのアドレスを入力として、該当するバッファへのファイル操作要求を禁止するロック機能161と、それを解除するアンロック機能162。

【0043】3. ファイル名、オフセット、データ長を入力として、該当するデータを格納しているディスクの位置（ディスクブロック）を返す。

【0044】ディスクコントローラ140は、ディスク要求処理手段142と、ディスク104から読み出した、あるいは、ディスク104へ書き込むデータをいったん格納するディスクキャッシュ141を備える。

【0045】サーバ計算機100が実行するネットワークカードドライバ170は、以下の機能を備える。

【0046】1. ネットワークカード130からの割り込みを処理する。

【0047】2. 通常のバケットの送受信を行う。

【0048】3. 主記憶120上に、送信バッファ122、受信バッファ123を確保する。

【0049】サーバ計算機100のプロセッサ101が実行するワーカーズレッド181は、リモートファイルシステム要求を受け、その処理を担当する。

【0050】リモートファイルシステム要求は、クライアント計算機200が、サーバ計算機100へ要求を送信するところから処理が始まる。以下、クライアント計算機200がリモートファイルシステム要求を出し、その要求をサーバ計算機100がどのように処理するかを図3のフローチャートにより説明する。

【0051】まず、ネットワークカード130のリモートファイル処理手段136は、クライアント計算機200からバケットの受信があると、以下の動作を行う。ここでは、通信にイーサネット、プロトコルにTCP/IPを用いている例を説明する。

【0052】ネットワークカード130宛てのバケットが、ネットワーク300上に発信された場合、そのバケットを受信バッファ135に蓄積する（ステップA1）。

【0053】受信バッファ135に蓄積したバケットを解析し、TCP/IPのプロトコル処理を行うバケット解析部1360がバケットのデータ部を解析し、処理できるプロトコルかどうか判定する（ステップA2）。

【0054】処理できないプロトコルの場合（ステップ

A2のNo)、ステップA8の処理を行う。

【0055】処理できるプロトコルの場合(ステップA2のYes)、パケット解析部1360によりプロトコル処理を行う(ステップA3)。

【0056】次に、リモートファイルシステム180への要求かどうかを判定する(ステップA4)。ここで、リモートファイルシステム180への要求は、あらかじめ定められたTCP/IPのポートに対し送信されるので、パケットのポート番号を監視すれば、容易に判定することができる。

【0057】リモートファイルシステム180への要求の場合は(ステップA4のYes)、ステータスレジスタ132を1(ファイルシステムのリードライト要求受信)にする(ステップA5)。

【0058】パケットからファイルシステム要求を取り出し、ファイル要求バッファ133に格納する。その要求は、ファイル操作、ファイル名、オフセット、データ長等からなる。さらに、要求がリードの場合は、ネットワークカード130上の送信バッファ134を確保して、そのアドレスをファイル要求バッファ133のデータアドレスに格納する。ライト要求の場合は、書き込むべきデータを格納したネットワークカード130上の受信バッファ135のアドレスをファイル要求バッファ133のデータアドレスに格納する(ステップA6)。

【0059】その後、プロセッサへ割り込みをかける(ステップA7)。

【0060】リモートファイルシステム180への要求でない場合、即ち、リード、ライト処理以外のファイルシステム要求の場合は(ステップA4のNo)、通常のパケットとして、ネットワークカードドライバ170があらかじめ確保した主記憶120上の受信バッファ135へパケットを転送する。ステータスレジスタを2(通常パケット受信)にして(ステップA8)、プロセッサへ割り込みをかける(ステップA7)。

【0061】データ転送が多い処理の場合は、本発明で提案する方法でデータをネットワークカード130、ディスクコントローラ140の間で直接やりとりする。本実施例の説明では、リード、ライト要求のみを扱っているが、その他にもデータ転送量が大きい要求があれば、同様に処理を行う。

【0062】次に、ネットワークカード130のDMA転送の処理動作を図4のフローチャートを参照して説明する。

【0063】転送要求レジスタ131にセット(格納)されている送信元アドレスから送信先アドレスへDMA転送でデータを送る(ステップB1)。

【0064】ステータスレジスタ132を3にする。3は、DMA転送完了を示す(ステップB2)。

【0065】プロセッサに割り込みをかける(ステップB3)。

【0066】次に、ネットワークカード130の送信の処理動作を図5のフローチャートにより説明する。

【0067】パケット解析部1360により、パケットが通常のパケット送信か否かを判定する(ステップC1)。

【0068】通常のパケット送信の場合(ステップC1のYes)、通常のパケット送信を実行する(ステップC2)。その後、ステータスレジスタ132を4にして(4は、通常のパケット送信処理を示す)(ステップC3)、プロセッサに割り込みをかける(ステップC4)。

【0069】通常のパケット送信でない場合(ステップC1のNo)、送信バッファのデータ134とファイル要求からクライアント計算機200への返答を作成する(ステップC5)。この作成したパケットについてパケット解析部1360によりプロトコル処理を行う(ステップC6)。パケットをネットワークに送信する(ステップC7)。

【0070】ネットワークカードドライバ170は、サーバ計算機100のプロセッサ101が実行する、ネットワークカード制御ソフトウェアである。その処理動作を図6のフローチャートにより説明する。

【0071】プロセッサが割り込みを受け、ネットワークカードドライバ170の割り込み処理ルーチンが呼び出される。

【0072】ネットワークカードドライバ割り込み処理ルーチンは、リモートファイルシステム180への要求か、通常のパケットかを、ネットワークカード130のステータスレジスタを調べることで判定する(ステップD1)。

【0073】リモートファイルシステム180への要求ではない場合、即ち、ステータスレジスタ132に格納された値が1以外の場合は(ステップD1のNo)、ネットワークカードドライバ170の割り込み処理ルーチンは、通常ネットワークカードドライバ170と同様の動作をする。

【0074】ここで、ステータスレジスタ132が2かどうかを判定する(ステップD4)。

【0075】ステータスレジスタ132に格納された値が2の場合(ステップD4のYes)、ネットワークカードドライバ170の割り込み処理ルーチンは、ネットワークカード130が転送したパケットのデータのうち、上位のプロトコル処理に不要なヘッダ(イーサネットフレーム情報)等を取り除き、プロトコルドライバ(TCP/IP等)に渡す(ステップD5)。以降は、通常のパケット受信と同じ処理になるので、ここでは説明を省略する。

【0076】リモートファイルシステム180への要求の場合、即ち、ステータスレジスタ132に格納された値が1の場合は(ステップD1のYes)、ネットワー

クカードドライバ170の割り込み処理ルーチンは、ネットワークカード130のファイル要求バッファ133から、ファイル要求を取り出す(ステップD2)。

【0077】このリモートファイルシステム要求を処理するワーカースレッド181を起動し、要求を渡す(ステップD3)。

【0078】ステータスレジスタ132に格納された値が2以外の場合(但し、1は、既に除外されている)(ステップD4のNo)、ステータスレジスタ132の値が3かどうかを判定する(ステップD6)。

【0079】ステータスレジスタ132の値が3の場合(ステップD6のYes)、DMA転送を要求しているスレッドを起動する(ステップD7)。

【0080】ステータスレジスタ132の値が3以外の場合(但し、1と2は既に除外されている)(ステップD6のNo)、ステータスレジスタ132の値が4かどうかを判定する(ステップD8)。

【0081】ステータスレジスタ132の値が4以外の場合(但し、1と2と3は既に除外されている)(ステップD8のNo)、処理は終了する。

【0082】ステータスレジスタ132の値が4の場合(ステップD8のYes)、通常のバケット送信完了処理を実行する(ステップD9)。

【0083】上記、図6のフローチャートの処理動作により、ネットワークカードドライバ170の割り込み処理はこれで完了し、以降、プロセッサが実行するワーカースレッド181(リモートファイルシステム処理を行う)が、次に説明する一連の処理動作を実行する。

【0084】この処理動作を図7のフローチャートを参照して説明する。

【0085】ワーカースレッド181は、渡されたファイル操作が、リードか、ライトか、その他の操作か、によって以下の処理を行う。

【0086】まず、ワーカースレッド181は、渡されたファイル操作が、リード要求か否かを判定する(ステップE1)。

【0087】リード要求の場合(ステップE1のYes)、ワーカースレッド181は、ファイル要求のファイル、オフセット、データ長から、リードしたい部分が主記憶120上のバッファキャッシュ121に存在するかどうかを、ローカルファイルシステム160の機能を使って調べる(ステップE2)。

【0088】バッファが存在する場合(ステップE2のYes)は、該当するバッファのアドレスを得ることができる。

【0089】バッファキャッシュ121に存在する場合は、以下の処理を行う。

【0090】該当するバッファをロックする操作を、ローカルファイルシステム160の機能を用いて行う(ステップE3)。この操作により、そのバッファに対する

新たな操作要求が来た場合には、その要求の処理を遅延する排他を行ったことになる。

【0091】ワーカースレッド181は、転送要求レジスタ131にリードしたいバッファキャッシュ121上のデータのアドレスと、このデータを格納するネットワークカード130の送信バッファ134上のアドレスをセット(格納)し、この格納されたアドレスに基づき、バッファキャッシュ121のバッファの内容をネットワークカード130へDMA転送する要求を、ネットワークカード130のリモートファイル処理手段136に出し、その完了を待つ(ステップE4)。

【0092】ネットワークカード130のリモートファイル処理手段は、上記DMA要求を受けて、ネットワークカード130上の送信バッファ134にバッファを用意し、渡された主記憶120上のバッファのアドレスが指すデータをDMA転送する(ステップE5)。

【0093】ネットワークカード130のリモートファイル処理手段136は、DMA転送が終了すると、ステータスレジスタ132を3(DMA転送完了)にして、プロセッサに割り込みをかける(ステップE6)。

【0094】ネットワークカードドライバ170の割り込みルーチンが呼び出される。割り込みルーチンは、DMA転送完了の割り込みであることをステータスレジスタ132を読むことで認識する。ステップE5の処理で待っているワーカースレッド181を起動し、割り込み処理を終了する(ステップE7)。

【0095】ワーカースレッド181は、バケット送信要求をネットワークカード130のリモートファイル処理手段136に出す(ステップE8)。

【0096】リモートファイルシステム要求(リード要求)に対してのデータが、(2-3)の転送でネットワークカード130上の送信バッファに入っているのを、後は、送信を実行すればよい。

【0097】ネットワークカード130のリモートファイル処理手段136は、送信バッファ134のデータを送信元(リード要求を出したクライアント計算機200)に送り返すためにイーサネットパケットのヘッダを作成し、送信を実行する(ステップE9)。

【0098】ワーカースレッド181は、ローカルファイルシステム160の機能を用いて、主記憶120のバッファキャッシュ121のバッファへのロックを解除する(ステップE10)。以後、該当するバッファへの操作を他のスレッド、プロセス等が行うことができる。

【0099】バッファに存在しない場合は、以下の操作を行う(ステップE2のNo)。

【0100】ワーカースレッド181は、ローカルファイルシステム160の機能を用いて、ファイルの読み出し操作で必要となるディスクブロックを調べる。ファイル名、オフセットの情報から、ディスク上のどの部分に要求されるデータが存在するかを調べる。ディスク10

4上の位置は、通常ディスクブロック番号等で管理されているので、ここではディスクブロック番号を得る（ステップE11）。

【0101】ワーカースレッド181は、ステップE11で得たディスクブロックを読み出し、ネットワークカードドライバ170の割り込みルーチンから渡されたネットワークカード130上の送信バッファ134にDMA転送を行うために、ディスクドライバ150を呼び出す（ステップE12）。

【0102】ディスクドライバ150は、その要求をディスクコントローラ140に伝え、完了を待つ（ステップE13）。

【0103】ディスクコントローラ140のディスク要求処理手段は、データをいったんディスク104から読み出し、そのデータをディスクキャッシュ141に格納する。そのキャッシュデータをステップE12で指示されたネットワークカード130の送信バッファ134に転送する（ステップE14）。

【0104】ディスクコントローラ140のディスク要求処理手段142は、プロセッサ101に割り込みをかける（ステップE15）。

【0105】ディスクドライバ150の割り込みルーチンは、ステップE13で、その完了を待っているワーカースレッド181を起動し、割り込み処理を終了する（ステップE16）。

【0106】ワーカースレッド181は、データ送信要求をネットワークカード130に出す（ステップE17）。

【0107】これは、クライアント計算機200から要求されたリモートファイルシステム要求（リード要求）のデータがネットワークカード130上に整ったので、ネットワークカード130に送信を指示することができるようになったからである。

【0108】ネットワークカード130のリモートファイル処理手段136は、送信バッファ134に入っているデータを使用して、リモートファイルシステム要求の返答を作成し、クライアント計算機200に送信する（ステップE18）。

【0109】ワーカースレッド181は、ローカルファイルシステム160の機能を用いて、ファイルの読み出し操作を出し、その完了を待つ（ステップE19）。

【0110】このファイル読み出し要求を処理する必要は、本来はない。なぜなら、クライアント計算機200から要求されたリード要求はすでに処理しているからである。しかし、ディスクから読み出したデータをサーバ計算機100のバッファキャッシュ121に格納しておき、次のアクセスがあった場合に、再利用できるようにしておくことで、キャッシュ効果が期待できるようにバッファキャッシュ121にデータを格納しておいてもよい。

【0111】その後、ワーカースレッド181の処理を終了する。

【0112】リード要求で無い場合（ステップE1のNo）、即ちライト要求の場合、ライト要求がライトスルー要求か否かを判定する（ステップE20）。

【0113】ライトスルー要求の場合（ステップE20のYes）、即ちディスクに即座にデータを書き込む要求の場合は、以下のような処理を行う。

【0114】ワーカースレッド181は、ファイル要求のファイル、オフセット、データ長から、ライトすべき部分が主記憶上のバッファキャッシュ121に存在するかどうかをローカルファイルシステム160の機能呼び出し調べる（ステップE21）。

【0115】ローカルファイルシステム160は、バッファが存在する場合（ステップE21のYes）、該当するバッファのアドレスを求め、ワーカースレッド181は、ローカルファイルシステム160の機能を用いて、そのバッファを破棄する（ステップE22）。これは、これから処理するライト要求によってデータが上書きされるので、主記憶120上のデータは、ディスク104に書き戻す必要がないためである。

【0116】ワーカースレッド181は、ローカルファイルシステム160の機能を用いて、ディスクブロックを調べ、ファイル要求のファイル、オフセット、データ長から、ライトすべきディスクブロックを得る（ステップE23）。

【0117】ワーカースレッド181は、ディスクドライバ150を用いて、ネットワークカード130の受信バッファ135上にあるデータをステップE23で得たディスクブロックへ書き込むように、要求を出し、その完了を待つ（ステップE24）。

【0118】ディスクコントローラ140のディスク要求処理手段142は、ディスク104の書き込みを終了すると、プロセッサ101に割り込みをかける（ステップE25）。

【0119】ディスクドライバ150の割り込み処理ルーチンは、ステップE24で完了を待っているワーカースレッド181を起こし、割り込み処理を終了する（ステップE26）。

【0120】ワーカースレッド181は、処理が完了したので、リモートファイルシステム要求に対する返答をクライアント計算機200に送信することを、ネットワークカード130に要求を出す（ステップE27）。

【0121】ネットワークカード130のリモートファイル処理手段136は、リモートファイルシステム要求の完了をクライアント計算機200に送信する（ステップE28）。

【0122】ワーカースレッド181は、ライト処理の指定を検査する（ステップE29）。

【0123】ライト処理がバッファキャッシュ121を

使用しないモードであれば（ステップE29のNo）、処理を終了する。

【0124】ライト処理がバッファを使用するモードの場合（ステップE29のYes）、バッファキャッシュ121にデータを入れておくために、ワークスレッド181は、ローカルファイルシステム160の機能を用いて、リード要求を処理し、その完了を待つ（ステップE30）。

【0125】このリード要求も本来は実行しなくてもよいものである。リモートファイルシステム要求を処理する場合、バッファキャッシュ121を使用する、しないは通常は、クライアント計算機200において意味があるものであるからである。

【0126】従って、ステップE30はなくてもよい。

【0127】その後、ワークスレッド181の処理を終了する。

【0128】ライト要求がライトスルーでない場合（ステップE20のNo）、以下の処理を行う。

【0129】ワークスレッド181は、ローカルファイルシステム160の機能を用い、ファイル要求のファイル、オフセット、データ長から、ライトすべき部分が主記憶上のバッファキャッシュ121に存在するかどうかを判定する（ステップE31）。

【0130】バッファが存在する場合（ステップE31のYes）、該当するバッファのアドレスを得ることができる。バッファが存在する場合は、以下の処理を行う。

【0131】ワークスレッド181は、ステップE31で得たバッファをロックする処理をローカルファイルシステム160の機能を用いて行う（ステップE32）。

【0132】ワークスレッド181は、受信バッファ135にあるデータのアドレスと、バッファキャッシュ121上のアドレスを転送要求レジスタ131にセットし、このアドレスに基づき、ネットワークカード130上の受信バッファ135にあるデータをステップE31で得たバッファキャッシュ121上のアドレスへDMA転送するように、ネットワークカード130のリモートファイル処理手段136に要求を出し、その完了を待つ（ステップE33）。

【0133】この転送により、バッファキャッシュ121上のデータを、リモートファイル要求（ライト要求）のデータに置き換えることになる。

【0134】ネットワークカード130のリモートファイル処理手段136は、DMA転送を完了したら、ステータスレジスタを3にセットして、プロセッサ101に割り込みをかける（ステップE34）。

【0135】ネットワークカードドライバ170の割り込みルーチンは、ステップE33で待つワークスレッド181を起こし、割り込みを完了する（ステップE3

5）。

【0136】ワークスレッド181、ロックしたバッファキャッシュ121上のバッファのロックを解除する（ステップE36）。

【0137】その後、ワークスレッド181の処理を終了する。

【0138】バッファがない場合（ステップE31のNo）、以下の処理を行う。

【0139】ワークスレッド181は、ローカルファイルシステム160の機能を用いて、バッファキャッシュ121上にバッファを確保し（ステップE37）、そのバッファをロックする（ステップE32）。

【0140】ステップE32以降は、バッファがある場合と同様の処理を続ける。

【0141】以上の様に、従来は、ネットワークカード130とディスクの間のデータ転送は、主記憶120にいったんデータを蓄積して行わなければならなかった。しかし、本発明によれば、リモートファイルシステム要求をネットワークカード130、ネットワークカードドライバ170、ディスクドライバ150とファイルシステムが協調して処理することで、ネットワークカード130とディスクコントローラ140の間で直接データをやりとりできる。また、そのデータのやりとりを主記憶120上のファイルシステムのバッファとの一貫性を保ちながら行っているの、サーバ計算機100上のローカルファイルシステム160がディスク104上のデータを操作する機能を維持している。

【0142】

【発明の効果】以上詳記したように本発明によれば、ネットワークカード、ディスクコントローラにあって、リモートファイル操作を処理するので、サーバ計算機のプロセッサは、リモートファイル操作を処理するためのプロトコル処理やファイル処理等を実行することがなくなり、それ以外の処理に資源、時間を割り当てることができる。

【0143】例えば、従来技術では、ライト処理の場合、ネットワークカードからデータを主記憶に転送し、更に主記憶からディスクへデータを転送するために、システムバス上に同一のデータが2回流れるものであったが、本発明によれば、ネットワークカードからディスクへI/Oバスのみを用いてデータを転送するので、システムバスを用いるデータの転送が必要ない。

【図面の簡単な説明】

【図1】本発明の一実施の形態に係わるネットワークファイルサーバシステムの構成を示すブロック図。

【図2】同実施の形態に係わるファイル要求バッファのデータ構成を示す図。

【図3】同実施の形態のネットワークカードの受信の際の処理動作を示すフローチャート。

【図4】同実施の形態に係わるネットワークカードのD



MA転送の際の処理動作を示すフローチャート。

【図5】同実施の形態に係わり、ネットワークカードの送信の際の処理動作を示すフローチャート。

【図6】同実施の形態に係わり、ネットワークカードドライバの割り込み処理ルーチンの処理動作を示すフローチャート。

【図7】同実施の形態に係わり、ワークスレッドの処理動作を示すフローチャート。

【図8】従来のネットワークファイルサーバシステムの構成を示す図。

【符号の説明】

10…サーバ計算機

11…プロセッサ

12…システムバス

13…PCIバス

14…主記憶

15…バッファキャッシュ

16…ディスクコントローラ

17…ディスク

18…ネットワークカード

19…受信バッファ

20…ネットワークドライバ

21…プロトコルスタック

22…リモートファイルシステム

23…ローカルファイルシステム

24…ネットワーク

25…クライアント計算機

100…サーバ計算機

101…プロセッサ

102…システムバス

103…I/Oバス

104…ディスク

120…主記憶

121…バッファキャッシュ

122…送信バッファ

123…受信バッファ

130…ネットワークカード

131…転送要求レジスタ

132…ステータスレジスタ

133…ファイル要求バッファ

1330…リモートファイル要求

1331…ファイル名

1332…オフセット

1333…データ長

1334…データアドレス

134…送信バッファ

135…受信バッファ、

136…リモートファイル処理手段

1360…パケット解析部

140…ディスクコントローラ

141…ディスクキャッシュ

142…ディスク要求処理手段

150…ディスクドライバ

160…ローカルファイルシステム

161…ロック機能

162…アンロック機能

170…ネットワークカードドライバ

180…リモートファイルシステム

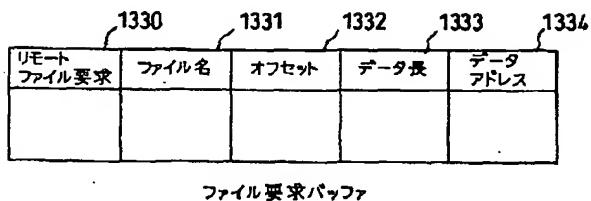
181…ワークスレッド

200…クライアント計算機

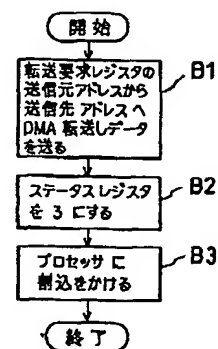
300…ネットワーク

1330…リモートファイル要求

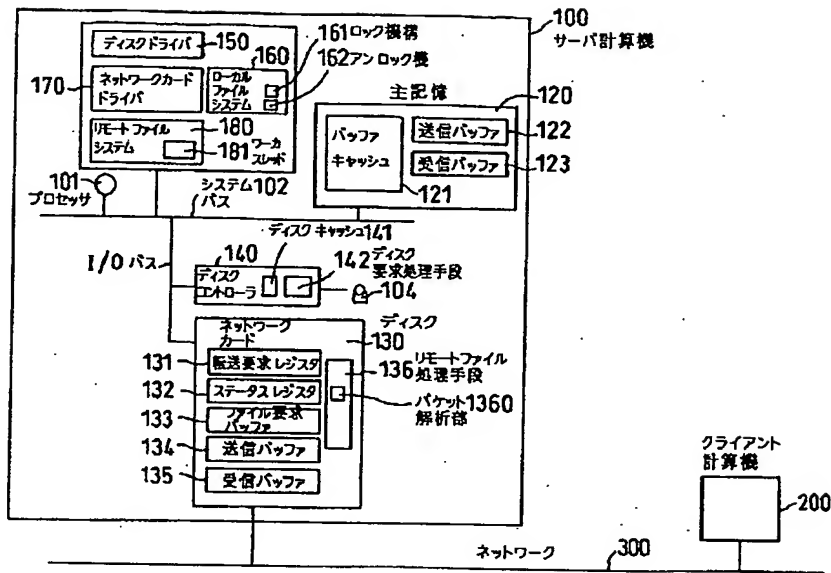
【図2】



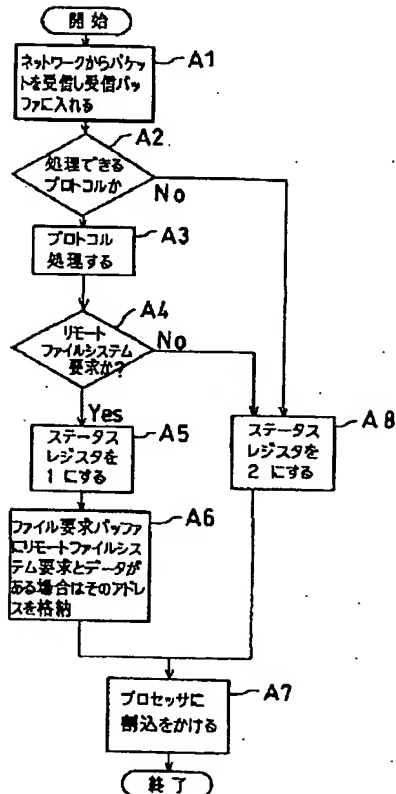
【図4】



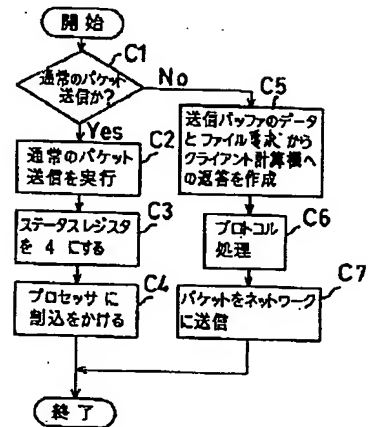
【図1】



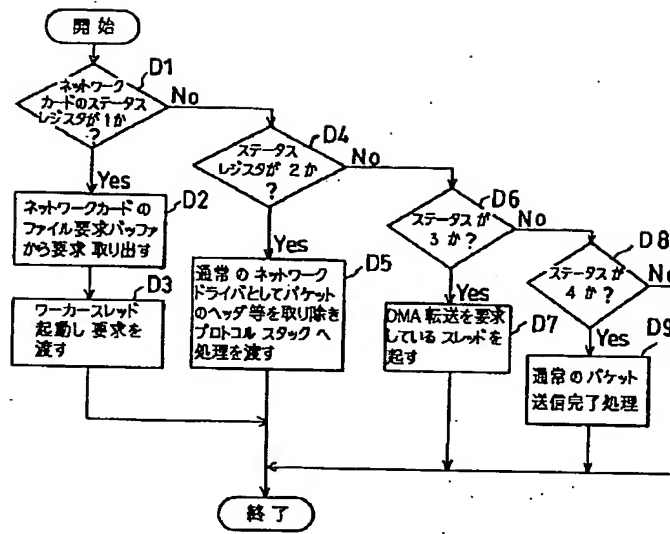
【図3】



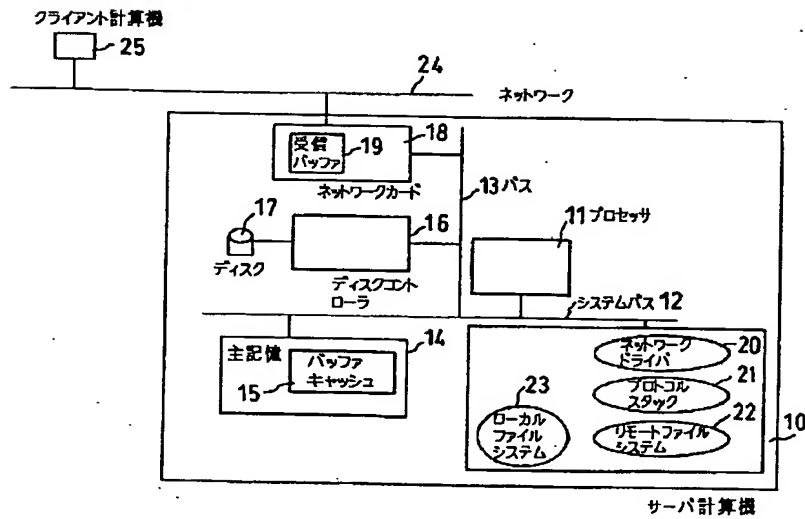
【図5】



【図6】



【図8】



【図7】

